# Applied Stochastic Processes, Chaos Modeling, and Probabilistic Properties of Numeration Systems

By Vincent Granville, Ph.D.

[MLTechniques.com](MLTechniques.com)

June 2018

This book is intended for professionals in data science, computer science, operations research, statistics, machine learning, big data, and mathematics. In 100 pages, it covers many new topics, offering a fresh perspective on the subject. It is accessible to practitioners with a two-year college-level exposure to statistics and probability. The compact and tutorial style, featuring many applications (Blockchain, quantum algorithms, HPC, random number generation, cryptography, Fintech, web crawling, statistical testing) with numerous illustrations, is aimed at practitioners, researchers and executives in various quantitative fields.

New ideas, advanced topics, and state-of-the-art research are discussed in simple English, without using jargon or arcane theory. It unifies topics that are usually part of different fields (data science, operations research, dynamical systems, computer science, number theory, probability) broadening the knowledge and interest of the reader in ways that are not found in any other book. This short book contains a large amount of condensed material that would typically be covered in 500 pages in traditional publications. Thanks to cross-references and redundancy, the chapters can be read independently, in random order.

This book is available for Data Science Central members exclusively. The text in blue consists of clickable links to provide the reader with additional references.  Source code and Excel spreadsheets summarizing computations, are also accessible as hyperlinks for easy copy-and-paste or replication purposes.

## About the author

Vincent Granville is a start-up entrepreneur, patent owner, author, investor, pioneering data scientist with 30 years of corporate experience in companies small and large (eBay, Microsoft, NBC, Wells Fargo, Visa, CNET) and a former VC-funded executive, with a strong academic and research background including Cambridge University.

## Content

The book covers the following topics:

### 1. Introduction to Stochastic Processes

We introduce these processes, used routinely by Wall Street quants, with a simple approach consisting of rescaling random  walks to make them time-continuous, with a finite variance, based on the central limit theorem.

- Construction of Time-Continuous Stochastic Processes
- From Random Walks to Brownian Motion

- Stationarity, Ergodicity, Fractal Behavior
- Memory-less or Markov Property
- Non-Brownian Process

## 2. Integration, Differentiation, Moving Averages

We introduce more advanced concepts about stochastic processes. Yet we make these concepts easy to understand even to the non-expert. This is a follow-up to Chapter 1.

- Integrated, Moving Average and Differential Process
- Proper Re-scaling and Variance Computation
- Application to Number Theory Problem

## 3. Self-Correcting Random Walks

We investigate here a breed of stochastic processes that are different from the Brownian motion, yet are better models in many contexts, including Fintech.

- Controlled or Constrained Random Walks
- Link to Mixture Distributions and Clustering
- First Glimpse of Stochastic Integral Equations
- Link to Wiener Processes, Application to Fintech
- Potential Areas for Research
- Non-stochastic Case

## 4. Stochastic Processes and Tests of Randomness

In this transition chapter, we introduce a different type of stochastic process, with number theory and cryptography applications, analyzing statistical properties of numeration systems along the way -- a recurrent theme in the next chapters, offering many research opportunities and applications. While we are dealing with deterministic sequences here, they behave very much like stochastic processes, and are treated as such. Statistical testing is central to this chapter, introducing tests that will be also used in the last chapters.

- Gap Distribution in Pseudo-Random Digits
- Statistical Testing and Geometric Distribution
- Algorithm to Compute Gaps
- Another Application to Number Theory Problem
- Counter-Example: Failing the Gap Test

## 5. Hierarchical Processes

We start discussing random number generation, and numerical and computational issues in simulations, applied to an original type of stochastic process. This will become a recurring theme in the next chapters, as it applies to many other processes.
- Graph Theory and Network Processes

- The Six Degrees of Separation Problem
- Programming Languages Failing to Produce Randomness in Simulations
- How to Identify and Fix the Previous Issue
- Application to Web Crawling

## 6. Introduction to Chaotic Systems

While typically studied in the context of dynamical systems, the logistic map can be viewed as a stochastic process, with an equilibrium distribution and probabilistic properties, just like numeration systems (next chapters) and processes introduced in the first four chapters.

- Logistic Map and Fractals
- Simulation: Flaws in Popular Random Number Generators
- Quantum Algorithms

## 7. Chaos, Logistic Map and Related Processes

We study processes related to the logistic map, including a special logistic map discussed here for the first time, with a simple equilibrium distribution. This chapter offers a transition between chapter 6, and the next chapters on numeration system (the logistic map being one of them.)

- General Framework
- Equilibrium Distribution and Stochastic Integral Equation
- Examples of Chaotic Sequences
- Discrete, Continuous Sequences and Generalizations
- Special Logistic Map
- Auto-regressive Time Series
- Literature
- Source Code with Big Number Library
- Solving the Stochastic Integral Equation: Example

## 8. Numerical and Computational Issues

These issues have been mentioned in chapter 7, and also appear in chapters 9, 10 and 11. Here we take a deeper dive and offer solutions, using high precision computing with BigNumber libraries.

- Precision Issues when Simulating, Modeling, and Analyzing Chaotic Processes
- When Precision Matters, and when it does not
- High Precision Computing (HPC)
- Benchmarking HPC Solutions
- How to Assess the Accuracy of your Simulation Tool

## 9. Digits of Pi, Randomness, and Stochastic Processes

Deep mathematical and data science research (including a result about the randomness of $\pi$, which is just a particular case) are presented here, without using arcane terminology or complicated

equations. Numeration systems discussed here are a particular case of deterministic sequences behaving just like the stochastic process investigated earlier, in particular the logistic map, which is a particular case.

- Application: Random Number Generation
- Chaotic Sequences Representing Numbers
- Data Science and Mathematical Engineering
- Numbers in Base 2, 10, 3/2 or $\pi$
- Nested Square Roots and Logistic Map
- About the Randomness of the Digits of $\pi$
- The Digits of $\pi$ are Randomly Distributed in the Logistic Map System
- Paths to Proving Randomness in the Decimal System
- Connection with Brownian Motions
- Randomness and the Bad Seeds Paradox
- Application to Cryptography, Financial Markets, Blockchain, and HPC
- Digits of $\pi$ in Base $\pi$

## 10. Numeration Systems in One Picture

Here you will find a summary of much of the material previously covered on chaotic systems, in the context of numeration systems (in particular, chapters 7 and 9.)

- Summary Table: Equilibrium Distribution, Properties
- Reverse-engineering Number Representation Systems
- Application to Cryptography

## 11. Numeration Systems: More Statistical Tests and Applications

In addition to featuring new research results and building on the previous chapters, the topics discussed here offer a great sandbox for data scientists and mathematicians.

- Components of Number Representation Systems
- General Properties of these Systems
- Examples of Number Representation Systems
- Examples of Patterns in Digits Distribution
- Defects found in the Logistic Map System
- Test of Uniformity
- New Numeration System with no Bad Seed
- Holes, Autocorrelations, and Entropy (Information Theory)
- Towards a more General, Better, Hybrid System
- Faulty Digits, Ergodicity, and High Precision Computing
- Finding the Equilibrium Distribution with the Percentile Test
- Central Limit Theorem, Random Walks, Brownian Motions, Stock Market Modeling
- Data Set and Excel Computations

## 12. The Central Limit Theorem Revisited

The central limit theorem explains the convergence of discrete stochastic processes to Brownian motions, and has been cited a few times in this book. Here we also explore a version that applies to deterministic sequences. Such sequences and treated as stochastic processes in this book.
- A Special Case of the Central Limit Theorem
- Simulations, Testing, and Conclusions
- Generalizations
- Source Code

## 13. How to Detect if Numbers are Random or Not

We explore here some deterministic sequences of numbers, behaving like stochastic processes or chaotic systems, together with another interesting application of the central limit theorem.

- Central Limit Theorem for Non-Random Variables
- Testing Randomness: Max Gap, Auto-Correlations and More
- Potential Research Areas
- Generalization to Higher Dimensions

## 14. Arrival Time of Extreme Events in Time Series

Time series, as discussed in the first chapters, are also stochastic processes. Here we discuss a topic rarely investigated in the literature: the arrival times, as opposed to the extreme values (a classic topic), associated with extreme events in time series.

- Simulations
- Theoretical Distribution of Records over Time

## 15. Miscellaneous Topics

We investigate topics related to time series as well as other popular stochastic processes such as spatial processes.

- How and Why: Decorrelate Time Series
- A Weird Stochastic-Like, Chaotic Sequence
- Stochastic Geometry, Spatial Processes, Random Circles: Coverage Problem
- Additional Reading (Including Twin Points in Point Processes)

## 16. Exercises